

Peer-to-Peer monitoring

P2Pmon

Design document for the P2Pmon-prototype
Developed by Aperte in cooperation with E-Secure-IT

Classified information removed – for redistribution
Version 1.0

Introduction

Over the last decade a number of peer-to-peer file-sharing networks have been in existence. The original goal of these networks (Napster, Kazaa) has been to facilitate the exchange of music between users. Due to legal prosecution, these two networks have been disabled or severely reduced in size.

There are currently other peer-to-peer networks however. Currently the major network is called Gnutella. It is a completely distributed network with a large number of clients (Limewire, Bearshare) and users. There has also been a shift in focus from users in these networks. Initially users were only exchanging music, but with the rise of broadband home internet connections the distribution of movies and other large files was also facilitated.

Music and movies aren't the only types of files that are distributed. As a side-effect, users can easily allow other users in the peer-to-peer network to access documents, spreadsheets and email archives, for instance. Often this is done without the knowledge of the end-user or with the end-user not considering the drawbacks of sharing his 'My Documents' folder. Confidential documents have been known to be leaked to the outside world via peer-to-peer networks, often with devastating effects.

In order to combat this type of document leakage, Aperte, together with E-Secure-IT, has developed a peer-to-peer monitoring system. This document details what the system is.

Key features of the peer-to-peer monitoring prototype are:

- 24/7 distributed search using multiple harvest nodes
- One single storage node with a web-based user interface
- Keyword-based search
- Filetype filters
- Host-specific searches
- Automatic summarization of the contents of files
- Details of the user who offer the files to be downloaded
- Flexible classification of document: severity, confidentiality and document type
- RSS feeds for easy integration

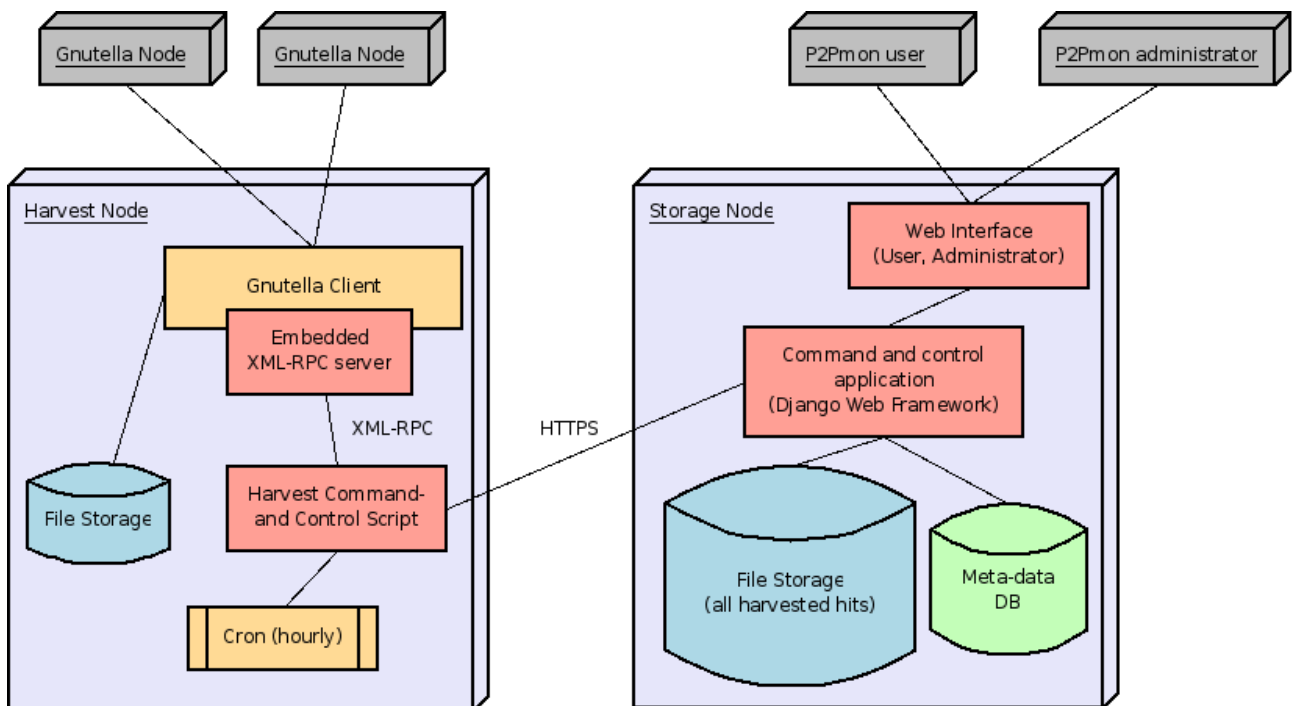
Design overview

The design of the peer-to-peer monitoring system is, like the network it monitors, distributed in nature. Thanks to the developed system, users can easily add and remove searches for files related to their business, thereby detecting document leakage.

A single storage node is active where all the retrieved files are selected and where users of the monitoring system can administer the network. Each file is automatically summarized upon retrieval at the storage node and the administrator interface offers means to classify each file 'hit'.

Multiple harvest nodes are used in order to retrieve as many results as possible. Currently the only harvest nodes available use a modified Gnutella client for searching and retrieving files. At fixed intervals the client sends out a search query for the keywords specified via the administrator interface and at fixed intervals the harvester node sends back the files found to the storage node for processing and collection.

An overview of the initial system



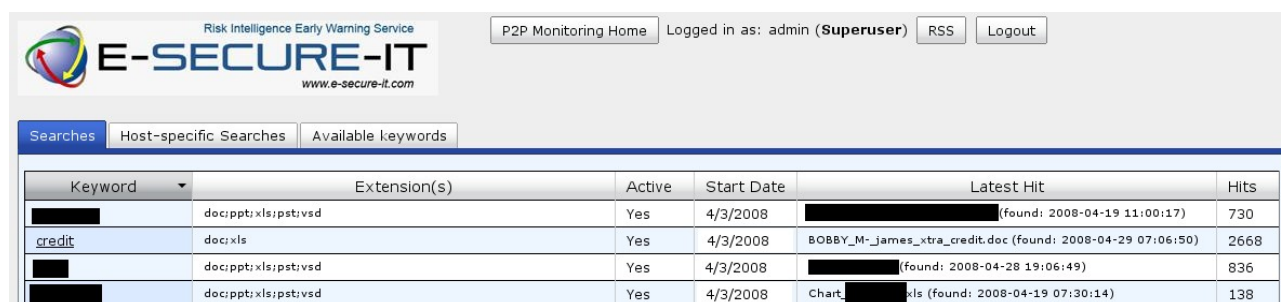
The above system is currently online and being tested for specific use-cases in the financial sector.

The current monitoring system

The current system has three main views: A view containing all the active searches, a view containing the results for each particular search and a view for retrieving and classifying a single hit. Via these views a user can filter and retrieve all results found by the system.

Due to the sensitive nature of the search queries performed the following paragraphs have been redacted.

Active-searches view:



The screenshot shows the E-SECURE-IT Risk Intelligence Early Warning Service interface. The header includes the logo, the text "Risk Intelligence Early Warning Service", and the URL "www.e-secure-it.com". There are navigation links for "P2P Monitoring Home", "Logged in as: admin (Superuser)", "RSS", and "Logout". Below the header, there are tabs for "Searches", "Host-specific Searches", and "Available keywords". The main content area displays a table of active searches.

Keyword	Extension(s)	Active	Start Date	Latest Hit	Hits
[REDACTED]	doc;ppt;xls;pst;vsd	Yes	4/3/2008	[REDACTED] (found: 2008-04-19 11:00:17)	730
credit	doc;xls	Yes	4/3/2008	BOBBY_M-james_xtra_credit.doc (found: 2008-04-29 07:06:50)	2668
[REDACTED]	doc;ppt;xls;pst;vsd	Yes	4/3/2008	[REDACTED] (found: 2008-04-28 19:06:49)	836
[REDACTED]	doc;ppt;xls;pst;vsd	Yes	4/3/2008	Chart, [REDACTED].xls (found: 2008-04-19 07:30:14)	138

This view allows the end-user to view the searches he/she has access to. It states the types of files being searched for, the start-date of the search and the latest file 'hit' that has been retrieved. This display is used as the main dashboard for the user, with the latest results available in a single overview.

Result list view:

The screenshot shows the E-SECURE-IT interface with the following elements:

- Header: Risk Intelligence Early Warning Service, E-SECURE-IT logo, P2P Monitoring Home, Logged in as: admin (Superuser), RSS, Logout.
- Search results: 730 hits for search: [redacted]. Showing page 2 from 49.
- Navigation: << previous page, all hits, next page >>, Apply fast-classification.
- Table of results:

ID	File name	KB	Host/Vendor	Class	L	D	FC
8229	[redacted]	21	202.84.125.28:6346 using BearShare Pro 5.2.5.3 from Philippines			3/18/2008	[dropdown]
	[redacted] Fraud Department [redacted].doc						

This view shows a paginated-list of results found for a single search keyword. Each result shows the filename and a short summary of the contents of the file. It also shows which host supplied the file and on which date the file was harvested.

Hit view:

The screenshot shows the E-SECURE-IT interface with the following elements:

- Header: Risk Intelligence Early Warning Service, E-SECURE-IT logo, P2P Monitoring Home, Search: citibank, Logged in as: admin (Superuser), RSS, Logout.
- Search hit details:
 - File: [redacted] (21504 B)
 - Keyword: [redacted]
 - Found on hosts: [dropdown] 202.84.125.28 [dropdown] Host hits
 - Found on: 2008-03-18 08:00:12 using hash: [redacted]
- File summary: [redacted]
- Classification: Submit classification
- Classification form:

Owner type	P2PMon Subscriber Customer [dropdown]	Country	South East Asia [dropdown]
File type	Document [dropdown]	Document type	Communication on Account [dropdown]
File dated	[dropdown]	Confidentiality level	Personal Confidential [dropdown]
Confidentiality details	Creditcard details [dropdown]	Criticality level	YELLOW [dropdown]

This view shows the complete details of a single hit, together with the complete automatically generated summary. By selecting the file, the end-user is able to retrieve the file for further manual processing. This view also offers a means of classifying the file, so allowing automatic notification for particular types of files and in the future automatic classification using the pre-classified files.

It is possible that a single file is offered by multiple hosts. This view shows a drop-down list of hosts that have done so. By selecting a single host, a host-specific result view is shown with all the files available from that host. This view also allows the end-user to specify a host-specific search (second tab on the Active-searches view). This type of search targets a single host and retrieves all files with a particular file-type.

The future of search in security

The need for knowing about certain activities in the area of IT security is one broader than file leakage on peer-to-peer networks. While the peer-to-peer monitoring prototype is being tested and further developed we also have the opportunity to look at other types of search on the internet.

The following types of search have been identified for future expansion:

- Web search, where existing search engines are used in order to gather changes on specific keywords
- RSS search, where a large number of feeds are automatically processed in order to determine the severity
- Usenet search, where specific security-related groups are to be targetted and processed.

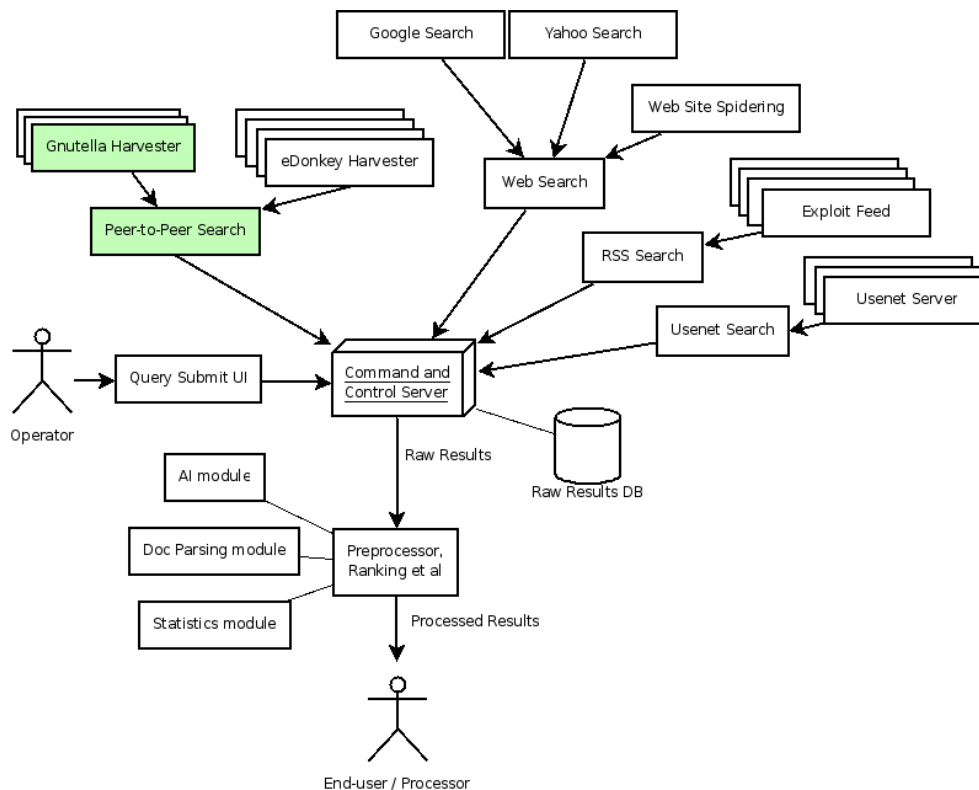
The idea is that a single operator submits a query, which is then used to search different types of internet-based content periodically for specific keywords and changes on the pages that were previously found.

The main problem is not in collecting this information, but in processing it. A number of strategies have been developed in order to preprocess the information:

- Document parsing (automatic summary, file meta-data)
- Statistics (comparing the number of times a keyword is used, if other keywords are used)
- AI-techniques (bayesian learning, for classification using pre-classified data)

The ideal result would be that the end-user gets a document every day detailing the most critical changes for the keywords he/she is registered to, with links to the files for further manual investigation.

An overview of our multi-headed approach:



Conclusion

With the current peer-to-peer monitoring prototype available, Aperte and E-Secure-IT have a stable system on which to build further offerings in the area of IT security and business intelligence. As we develop better classification methods and harvesters new improvements will become available to increase the accuracy and range of the system.

For further information, please contact E-Secure-IT (www.e-secure-it.com) or Aperte (www.aperte-it.com).